JOURNAL OF COMPLEXITY 5, 119–143 (1989)

Statistical Complexity of the Power Method for Markov Chains

PAUL E. WRIGHT

AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974

Received January 1987

Recently, attention has been focused on the statistical behavior of some of the classical algorithms of numerical and combinatorial analysis. For recent examples, see the work of Smale and others cited in the references. In his thesis, E. Kostlan (1985, "Statistical Complexity of Numerical Linear Algebra," Thesis, University of California, Berkeley) showed that the average time for convergence of the *squaring* algorithm of numerical analysis for finding dominant ε-eigenvectors of $n \times n$ real symmetric and Hermitian matrices is $O(\log(n - \log \varepsilon))$, $0 < \varepsilon < 1$, the average being taken over Gaussian ensembles of such matrices. We prove a complementary result for ensembles of $n \times n$ stochastic matrices, to the effect that for a large class of measures, $(1 + \delta) \log(n) + \log(-\log \varepsilon) + O(1)$ iterations suffice with probability $>1 - n^{-\delta}$, where δ is an arbitrary positive constant. This result has a direct translation which says that with asymptotically rare exceptions, Markov chains of size n require roughly at most $O(n^2)$ steps to reach equilibrium as n tends to ∞. 1989 Academic Press, Inc.

1. INTRODUCTION

Recently, attention has been focused on the statistical computational complexity of some of the classical algorithms of numerical and combinatorial analysis. Notable recent examples are the work of Smale (1981, 1983a,b, 1985) and Shub and Smale (1985, 1986) on Newton Iteration, generally convergent purely iterative algorithms for finding zeros of polynomials, and the Simplex method, as well as the work of Kostlan (1985), on numerical linear algebra. Further references to related recent work may be found in (Smale, 1985). In (Kostlan, 1985), the average number of iterations of the *squaring* algorithm for computing dominant ε-eigenvec-

0885-064X/89 $3.00

tors of $n \times n$ real symmetric and Hermitian matrices was found to be $O(\log(n - \log \varepsilon))$, $0 < \varepsilon < 1$. The ensembles over which averages are computed are the classical Gaussian ensembles (see, e.g., Mehta, 1967). We consider the performance of the same algorithm on statistical ensembles of row-stochastic matrices, applied to computing the dominant left-eigenvector to within an $\mathbf{L}^1$-accuracy of ε. (Since the matrices are row stochastic, the dominant right-eigenvector is the column-vector all of whose entries are unity.) Our results are to the effect that for a large class of probability distributions on the entries of the matrices in question, $(2 + \delta) \log(n) + \log(-\log \varepsilon) + O(1)$ iterations suffice, with $0 < \varepsilon < 1$ probability tending rapidly to 1 as n tends to ∞.

A strong case can be made for the probabilistic analysis of numerical algorithms, in which a measure is placed on the space of problems (i.e., input) and questions of computational complexity are investigated. In many problems such as the one at hand, one may even have a guaranteed rate of convergence of the sequence of iterates generated by the procedure, to a desired solution, for each problem instance. However, the rate of convergence often depends on the input, so that while for a specific instance of the problem one has, say, linear or quadratic convergence, the worst case running time to get within an accuracy of ε of the answer, where ε is some small tolerance, is actually unbounded. Naturally, expected running times do not tell the whole story, but on occasion (as we believe is the case here), it is the only way to involve the *size* of the problem in a running time estimate. Here, we look at the tail of the distribution of the running time. The *Achilles' heel* of this sort of analysis however, lies in the choice (or lack thereof) of a canonical measure on the space of problem instances. Here one may rely on various desired invariance properties of the measure (see, e.g., Kostlan's remarks on the Gaussian ensemble) to arrive at a reasonable or unique choice. If no canonical choice of measure can be found, one is then obliged to exhibit a class of measures containing "interesting ones," for which the algorithm in question may be shown to have (or lack) certain desirable properties, if this sort of analysis is to be of any merit. Our remarks here are in the spirit of this latter approach.

Stochastic matrices occur in many of the applications of mathematics to physical problems, as do the familiar real symmetric and Hermitian matrices, and are of interest and importance in their own right. It is well known (consult any standard text on probability theory) that the powers $\mathbf{P}^t$, $t = 1, 2, \ldots$, etc., of an irreducible stochastic matrix $\mathbf{P}$ converge (in various senses) to a rank one matrix with identical rows. The row-vector representing each row is the dominant left-eigenvector of $\mathbf{P}$, referred to as the stationary vector of $\mathbf{P}$. One interpretation of our results is that a Markov chain of size n requires roughly at most $O(n^2)$ steps to reach equilibrium, with asymptotically few exceptions as n tends to ∞. However, in this regard, other techniques are better adapted to the special structure of specific random walks. For example, in the case of random walks on groups, see the work of Aldous (1982, 1983). In these examples and others the matrices $\mathbf{P}$ which arise are doubly stochastic, so the dominant left-eigenvector is the row-vector of all ones and hence is known. There is also often a *sharp cutoff* phenomenon: distance of the powers of the stochasic matrix from their limiting form goes sharply from their initial value to a near zero, in an extremely narrow transition region. We make no structural assumptions about the matrix $\mathbf{P}$ save those which arise from natural statistical symmetry considerations. Our analysis then is roughly the study of the distribution of various quantities which control the asymptotic rate of convergence of the powers of the matrix to their limiting form.

Finally, we remark that for $\mathbf{P}$ as above, the classical Perron–Frobenius theory of nonnegative matrices implies the existence of a constant $0 < \lambda < 1$, such that the rate of convergence of $\mathbf{P}^t$ to its limiting form is governed by λ^t. For a certain class of stochastic matrices, Landau and Odlyzko (1981) demonstrate that $\lambda \leq 1 - n^{-3}$. A result of this type is of the sort that we seek and can be readily translated into a running time estimate for the *squaring* algorithm as well. The paper is organized as follows. Section 2 is devoted to general background material. The algorithm and the lemmas necessary to justify its correctness follow in Section 3. The Main Theorem and a few examples are the content of Section 4, while Section 5 is devoted to the proof of the Main Theorem. In Section 6, the class of probability densities associated with our ensembles is given an alternate characterization. This section is more technical than the rest, but the ideas contained therein are nonetheless fairly standard. Concluding remarks follow in Section 7. For the sake of clarity of exposition and ease of reading, the proofs of all lemmas or assertions peripheral to the proof of the Main Theorem have been relegated to the Appendix. Finally, as far as possible, we have attempted to adhere to standard mathematical notation. All vectors encountered in the text should be understood to be column-vectors. If π is a column-vector, π' denotes its transpose. Generally, a boldface quantity refers to either a vector or a matrix, or some other nonscalar, the meaning being clear from the context. Lg denotes the logarithm to the base 2, while log denotes the natural logarithm. The goal of this paper is the proof of the result stated below.

1.1. MAIN THEOREM. *For a large class of probability measures on the class of $n \times n$ stochastic matrices, the running time $T(k, \mathbf{P}, n)$ of the squaring algorithm for computing the dominant left-eigenvector to within an accuracy of 2^{-k} satisfies the inequality*

$$\Pr\{T(k, \mathbf{P}, n) > \lg(k + 2) + (2 + \delta) \lg(n) + K\} < \frac{1}{n^\delta}$$

for all $n > N$,

where δ is an arbitrary positive constant. The constant K above depends on the particular measure in question and may be explicitly given. The constant N also depends on the measure and is related to the behavior of the generating function of the measure near 1 (*see Section* 2).

Complete definitions and a full discussion may be found in Section 4.

2. General Background and Notation

Let $\mathbf{P}$ represent an $n \times n$ matrix $(\mathbf{P}_{ij})$, $1 \leq i, j \leq n$, of nonnegative reals. $\mathbf{P}$ is said to be (*row*) stochastic if its row sums are all equal to unity, i.e., $\sum_{j=0}^{n} \mathbf{P}_{ij} = 1$ for each i. $\mathbf{P}$ may be thought of as the transition matrix for a random walk on the edges of the complete digraph on n vertices, $K(n)$. $\mathbf{P}$ is said to be *irreducible* if the subgraph of $K(n)$ induced by the incidence matrix of $\mathbf{P}$ is strongly connected. $\mathbf{P}$ is further said to be of *period d* if for some (hence any) index i, $\gcd\{k > 0 : \mathbf{P}_{ii}^k > 0\} = d$. If $d = 1$, $\mathbf{P}$ is said to be *aperiodic*. The classical theorem in the theory of nonnegative matrices states:

2.1. Theorem (Perron-Frobenius). *Let $\mathbf{P}$ be an $n \times n$, irreducible, aperiodic stochastic matrix. Then there exists a unique strictly positive row-vector $\pi' = [\pi_1, \ldots, \pi_n]$, such that $\pi'\mathbf{P} = \pi'$. The eigenvalue* 1 *of $\mathbf{P}$ has algebraic* (*hence also geometric*) *multiplicity* 1, *and all other eigenvalues of $\mathbf{P}$ have absolute value strictly less than* 1.

2.2. Remark. (a) Without loss of generality, $\sum_{j=0}^{n} \pi_j = 1$ and this assumption is made hereinafter.

(b) There is no loss of generality from the point of view of applications in restricting our attention to aperiodic matrices. For, if $\mathbf{P}$ is irreducible, aperiodic or not, then for any $0 < \gamma < 1$, $\gamma\mathbf{P} + (1 - \gamma)\mathbf{I}$ (where $\mathbf{I}$ is the $n \times n$ identity matrix) is irreducible, aperiodic, and possesses the same dominant left-eigenvector as does $\mathbf{P}$.

(c) Let Σ^n denote the set of nonnegative row-vectors $\mathbf{x}' \in R^n$ such that $\sum_{j=0}^{n} x_j = 1$. Furthermore, let Π^n denote the n-fold direct product of $\Sigma^n \subset R^{n^2}$. Then Π^n may be identified with the space of $n \times n$ stochastic matrices. The subset of matrices $\mathbf{P}$ with all entries strictly positive (which are all aperiodic) forms a dense open set. For the distributions which we shall consider later, this subset will have full measure, so the reader may assume that we are always working with a matrix from this subset.

An excellent reference for the theory of nonnegative matrices is (Seneta, 1980). We shall strive however to keep our discussion as self-contained as possible. The following quantities will prove useful in our analysis.

2.3. DEFINITIONS. For $\mathbf{A} \in M_{n\times n}(R)$ the set of $n \times n$ real matrices, let

$$\|\mathbf{A}\| \equiv \max_{1\le i\le n} \sum_{j=1}^{n} |A_{ij}|. \tag{2.1}$$

For $\mathbf{x} \in R^n$ let,

$$\|\mathbf{x}\|_\infty \equiv \max_{1\le i\le n} |x_i|, \tag{2.2a}$$

$$\|\mathbf{x}\|_1 \equiv \sum_{i=1}^{n} |x_i|, \tag{2.2b}$$

and

$$\sigma(\mathbf{x}) \equiv \max_{1\le i\le n} x_i - \min_{1\le i\le n} x_i. \tag{2.3}$$

N.B. $\sigma(\mathbf{x}) = 0$ iff $\mathbf{x}' \propto \mathbf{1}'$, the row-vector consisting of all ones.

Finally, for $\mathbf{x} \in R^n$ let $\Lambda(\mathbf{x})$ denote the $n \times n$ real matrix with

$$(\Lambda(\mathbf{x}))_{ij} \equiv x_j, \qquad 1 \le i, j \le n. \tag{2.4}$$

We state the following easy proposition without proof.

2.4. PROPOSITION. (a) $\|\cdot\|$ *is a Banach norm on* $M_{n\times n}(R)$; *i.e., it is a norm in the usual sense and* $\|\mathbf{AB}\| \le \|\mathbf{A}\|\cdot\|\mathbf{B}\|$, *for* $\mathbf{A}, \mathbf{B} \in M_{n\times n}(R)$.

(b) *If* $\mathbf{P}$ *is stochastic then* $\|\mathbf{P}\| = 1$.

(c) *For* $\mathbf{x} \in R^n$, $\|\Lambda(\mathbf{x})\| = \|\mathbf{x}\|_1$.

(d) *For* $\mathbf{x} \in R^n$, *we have* $\|\mathbf{Ax}\|_\infty \le \|\mathbf{A}\|\, \|\mathbf{x}\|_\infty$ *and* $\|\mathbf{x}'\mathbf{A}\|_1 \le \|\mathbf{A}\|\, \|\mathbf{x}\|_1$.

The following lemma is standard, but it plays a key role in our analysis. It may be assembled from facts contained in (Seneta, 1980, Chap. 3), where $\tau(\mathbf{P})$ appears as $\tau_1(\mathbf{P})$. Its elementary proof is deferred to the Appendix.

2.5. DEFINITION. Let $\mathbf{A}$ be an $n \times n$ real matrix, and set

$$\tau(\mathbf{A}) \equiv \frac{1}{2} \max_{i\,\{j} \sum_{k=0}^{n} |A_{ik} - A_{jk}|. \tag{2.5a}$$

Clearly, for $\mathbf{A} = \mathbf{P}$ a stochastic matrix one has $0 \leq \tau(\mathbf{P}) \leq 1$ and, furthermore,

$$\tau(\mathbf{P}) = 1 - \min_{i<j} \sum_{k=0}^{n} \min (P_{ik}, P_{jk}). \tag{2.5b}$$

2.6. LEMMA. *Let* $\mathbf{P}$ *be an irreducible, aperiodic stochastic matrix with stationary vector* π'. *Then*

(a) $\frac{1}{2} \|\mathbf{P} - \Lambda(\pi)\| \leq \tau(\mathbf{P}) \leq \|\mathbf{P} - \Lambda(\pi)\|$,

(b) $\tau(\mathbf{P}) = \sup \{\sigma(\mathbf{Px})/\sigma(\mathbf{x}) : \mathbf{x} > 0, \sigma(\mathbf{x}) \neq \mathbf{0}\}$,

(c) $\tau(\mathbf{P}_1\mathbf{P}_2) \leq \tau(\mathbf{P}_1)\tau(\mathbf{P}_2)$, *and*

(d) $\|\mathbf{P}^t - \Lambda(\pi)\|$ *is a decreasing function of* $t = 1, 2, \ldots$, *etc.*

When the matrix $\mathbf{P}$ is understood, we write

$$d(t) \equiv \tfrac{1}{2}\|\mathbf{P}^t - \Lambda(\pi)\|, \tag{2.6a}$$

$$\rho(t) \equiv \tau(\mathbf{P}^t), \qquad t = 1, 2, \ldots. \tag{2.6b}$$

By translation one has immediately,

2.7. COROLLARY. *Let* $\mathbf{P}$ *be as above. Then*

(a) $d(t) \leq \rho(t) \leq 2d(t)$,

(b) $\rho(\cdot)$ *is submultiplicative, i.e.,* $\rho(s + t) \leq \rho(s)\rho(t)$, $s, t \geq 0$, *and*

(c) $d(\cdot)$ *is a decreasing function.*

The reader is invited to see (Aldous, 1982) for an interesting probabilistic proof of this last corollary, using *coupling* techniques, as well as an interpretation of these quantities in terms of the total variation norm on the set of probability measures on the set $\{1, \ldots, n\}$. The proofs in the Appendix, however, are completely algebraic in nature.

The final definition we need before proceeding is that of the convolution of measures. By a *probability measure* on $[0, \infty)$, we shall mean a finite positive *Borel* measure of total mass 1. Let μ and ν be probability measures on $[0, \infty)$. The convolution of μ and ν, denoted $\mu\nu$, is the unique measure on $[0, \infty)$ such that

$$\int_0^\infty f(\eta)\, d\mu\nu(\eta) = \int_0^\infty \int_0^\infty f(\zeta + \xi)\, d\mu(\zeta)\, d\nu(\xi), \tag{2.7}$$

for all bounded measurable functions $f(\cdot)$. For E a Borel set in $[0, \infty)$, we have

$$(\mu\nu)(E) = (\mu \times \nu)\ (\{(x, y) \in [0, \infty \times [0, \infty) \text{ s.t. } x + y \in E\}), \quad (2.8)$$

where $\mu \times \nu$ denotes the product measure on $[0, \infty) \times [0, \infty)$. If μ and ν are probability measures, then so is $\mu\nu$. The interested reader may consult (Rudin, 1984) for technical details. If μ is a probability measure on $[0, \infty)$, define

$$G_\mu(\xi) \equiv \int_0^\infty \xi^r \, d\mu(r), \quad (2.9)$$

where Δ is the open unit disk in the complex plane. By definition, ξ^r is equal to $e^{r \log \xi}$, using the branch of the logarithm in the complex plane minus the set $(-\infty, 0]$, giving a real determination of the logarithm on the positive real axis. G_μ is called the *generating function* of μ. The proof of the following easy proposition may be found in the Appendix.

2.8. PROPOSITION. *G_μ is analytic in the domain $\Delta \setminus (-1, 0]$, and one has there*

$$G_\mu^{(k)}(\xi) = \int_0^\infty r(r - 1)(r - 2) \ldots (r - k + 1)\xi^{(r-k)} \, d\mu(r). \quad (2.10)$$

3. THE ALGORITHM

The squaring algorithm is quite simple. Let $0 \leq \varepsilon \leq 1$ be given. $\mathbf{P}$ is an $n \times n$ stochastic matrix, and $\mathbf{x}$ is a real n-vector, which will contain the ε-approximate stationary vector of $\mathbf{P}$ on termination. We have the following algorithm.

Algorithm SQUARE ($\mathbf{P}$, $\mathbf{x}$, ε):

(1) $\mathbf{P} \leftarrow \mathbf{P}^2$.

(2) $x_j \leftarrow \frac{1}{n} \sum_{i=0}^{n} P_{ij}, j = 1, 2, \ldots, n.$

(3) **If** $(\|\mathbf{P} - \Lambda(\mathbf{x})\| > \varepsilon)$ **goto** (1).

(4) **Stop.**

The following lemma establishes the correctness of the above procedure. The proof may be found in the Appendix.

3.1. LEMMA. *Let* $\mathbf{x}$ *be as in the algorithm*:

(a) *If* $\|\mathbf{P} - \Lambda(\pi)\| \le \varepsilon$ *then* $\|\mathbf{P} - \Lambda(\mathbf{x})\| \le 2\varepsilon$ *and* $\|\mathbf{x} - \pi\|_1 \le \varepsilon$.

(b) *If* $\|\mathbf{P} - \Lambda(\mathbf{x})\| \le \varepsilon$ *then* $\|\mathbf{P} - \Lambda(\pi)\| \le 2\varepsilon$ *and* $\|\mathbf{x} - \pi\|_1 \le \varepsilon$.

In short, if $\mathbf{P}$ *is close to* $\Lambda(\pi)$, *the algorithm halts, and the converse is true. Upon termination,* $\mathbf{x}$ *is within an* L^1 *ball of radius* ε *about* π.

3.2. DEFINITION.

$$T(k, \mathbf{P}, n) = \min \left\{ \begin{matrix} t > 0: \textit{algorithm stops in } t \textit{ steps with} \\ n \times n\ \mathbf{P}, \textit{ and } \varepsilon = 2^{-k} \end{matrix} \right\}. \tag{3.1}$$

The above lemma together with Corollary 2.7 implies

$$T(k, \mathbf{P}, n) \le \min\{t : \tau^{2^t} \le 2^{-(k+2)}\}. \tag{3.2}$$

3.3. *Remark.* Since the powers of a stochastic matrix are again stochastic, the algorithm may be viewed as a (particularly simple) dynamical system on Π^n, whose fixed points are (by Lemma 2.6) precisely the matrices of the form $\Lambda(\mathbf{x})$ for some vector $\mathbf{x}$.

4. THE STATISTICAL ENSEMBLE

Since $\sum_{j=1}^n P_{ij} = 1$ for each i, we may think of the P_{ij} for fixed i as the interval lengths obtained by partitioning the unit interval $[0, 1]$ by $n - 1$ points chosen according to some distribution. In (Kostlan, 1985), the Gaussian ensembles of real symmetric or Hermitian matrices are the unique choices of statistical ensembles if one insists on a measure on the matrices which is invariant under the action of the orthogonal or unitary groups, respectively. This criterion arises from desire to have the results independent of the manner in which the axes in Euclidean space which give rise to the matrix elements were chosen. Similarly, here we shall insist on the invariance of the distribution of the stochastic matrices $\mathbf{P}$, under permutation of the labels of the sites of the associated random walks they describe. This implies in particular that all off-diagonal elements of the random matrix $\mathbf{P}$ from our ensemble should be identically distributed, as should the diagonal elements. In addition, we assume the sites to be independent. A simple model satisfying the above requirements is

$$P_{ij} = \frac{X_{ij}}{\sum_{k=1}^n X_{ik}}, \qquad 1 \le i, j \le n, \tag{4.1}$$

where $\mathbf{X} = (X_{ij})$ is a matrix of nonnegative i.i.d. random variables. The common distribution of the elements of $\mathbf{X}$ will be assumed to have a density with respect to Lebesgue measure on $[0, \infty)$, so that with probability one, each entry is strictly positive. As a simple example, if the X_{ij} have the exponential density $\alpha \exp(-\alpha x)$, $x \geq 0$ ($\alpha > 0$ *fixed*), the P_{ij} are distributed as the interval lengths obtained by marks placed *uniformly* on the unit interval. See (Feller, 1966) for details. This example and the others which we shall consider are all related to the family of *gamma* densities,

$$f_{\alpha,r}(x) = \frac{\alpha^r x^{r-1}}{\Gamma(r)} \exp(-\alpha x), \qquad x \geq 0,\ \alpha,\ r > 0 \textit{ fixed}. \tag{4.2}$$

Here, $\Gamma(\cdot)$ denotes the usual gamma function. An important property of this family is its closure under convolution,

$$f_{\alpha,r} * f_{\alpha,s}(\cdot) = f_{\alpha,r+s}(\cdot), \qquad r,\ s > 0 \textit{ fixed}, \tag{4.3}$$

where for f, $g \in L^1([0, \infty))$,

$$(f * g)(x) \equiv \int_0^x f(x - y)g(y)\, dy, \qquad x \geq 0. \tag{4.4}$$

4.1. Definition. Let F_α denote the class of probability densities on $[0, \infty)$ of the form

$$f(x) = \int_0^\infty \frac{(\alpha x)^r}{\Gamma(r + 1)} \alpha \exp(-\alpha x)\, d\mu(r), \qquad x \geq 0,\ \alpha > 0 \textit{ fixed}, \tag{4.5}$$

where μ is a probability measure on $[0, \infty)$ with finite mean $\bar{\mu}$, i.e.,

$$\bar{\mu} = \int_0^\infty r\, d\mu(r) < \infty. \tag{4.6}$$

Elements of F_α may be thought of as mixtures of gamma densities. In Section 6, we give a simple proof of the often overlooked fact that the union of the F_α for positive α is dense in the unit ball of the positive cone of $L^1[0, \infty)$, i.e., the class of probability densities on $[0, \infty)$. The integral in (4.5) is readily seen to be finite by appealing to *Stirling's* formula (see, for example, Whittaker and Watson, 1962, p. 253),

$$\log \Gamma(r + 1) = \left(r + \frac{1}{2}\right) \log (r + 1) - (r + 1) + \frac{1}{2} \log (2\pi) + \frac{\theta(r)}{12(r + 1)}, \tag{4.7}$$

where $0 < \theta(r) < 1$, $r > 0$. Fubini's theorem (see, e.g., Rudin, 1974) shows that indeed

$$\int_0^\infty f(x)\, dx = 1 \qquad \text{for } \textit{all } f \in F_\alpha.$$

Examples of F_α are given below. Our result on running times is the content of the

4.2. MAIN THEOREM. *Let X_{ij} have density $f \in F_\alpha$, and let P_{ij} be defined by* (4.1). *Then there exist constants $0 \le K$, $N < \infty$ (dependent on f), such that for every $\delta > 0$,*

$$\Pr\{T(k, \mathbf{P}, n) > \lg(k + 2) + (2 + \delta) \lg(n) + K\} < \frac{1}{n^\delta}. \tag{4.8}$$

4.3. *Remark.* A similar result holds when f is the χ-squared density with one degree of freedom, but it is slightly weaker in that the lg (n) term above is replaced by 2 lg (n). However, in Section 6, we give a simple proof of the often overlooked fact that the union of the F_α for positive α is dense in the unit ball of the positive cone of $L^1[0, \infty)$, i.e., the class of probability densities on $[0, \infty)$.

4.4. EXAMPLES. In the examples below, μ has discrete support. The values of μ at its points of positive mass are given.

(a) *Gamma densities.* Here we take

$$\mu(E) = \begin{cases} 1, & \text{if } k \in E, \\ 0, & \text{otherwise} \end{cases} \tag{4.9}$$

with $k \ge 0$ fixed. The choices, $k = 0$, $k = m/2$, and $k = m$, for m a fixed positive integer yield the *exponential* density, the *χ-squared density with* $(m + 1)$ *degrees of freedom*, i.e., the sum of $m + 1$ *i.i.d. normal* random variates, and the *m-Erlang* density, i.e., the sum of m *i.i.d. exponential* random variates, respectively.

(b) *Bessel densities.* The modified Bessel function of order $\rho > 0$ may be defined by

$$I_\rho(x) = \sum_{k=0}^\infty \frac{1}{k!\Gamma(k + \rho + 1)} \left(\frac{x}{2}\right)^{2k+\rho}, \qquad x \ge 0. \tag{4.10}$$

Let μ be supported on the set $\{2k + \rho - 1 : k \textit{ a positive integer}\}$,

$$\mu_{2k+\rho-1} = \frac{\rho}{2^{2k+\rho}} \frac{\Gamma(2k+\rho)}{\Gamma(k+\rho+1)} \frac{1}{k!}. \tag{4.11}$$

The reader may easily confirm that then

$$f(x) = \frac{\rho}{x} e^{-\alpha x} I_\rho(\alpha x). \tag{4.12}$$

For integral ρ, the above densities arise in connection with the distribution of *first-passage* times through $\rho \geq 1$ for symmetric random walks on the integers with exponentially distributed holding times at each site. That (4.11) defines a genuine probability density on $[0, \infty)$ is a short exercise with the *hypergeometric* function left to the Appendix.

Another example of this type is obtained by choosing the support of μ to be the set $\{k + \rho,\ k \textit{ a positive integer}\}$ and setting with $\rho > 0$,

$$\mu_{k+\rho} = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \lambda > 0. \tag{4.13}$$

Then one obtains

$$f(x) = \left(\frac{x}{\lambda}\right)^{\rho/2} e^{-(\lambda+x)} I_\rho(2\sqrt{\lambda x}), \tag{4.14}$$

another well-known density, which differs from the previous one in its asymptotic behavior as x tends to ∞.

An alternate characterization of the densities in the class F_α is the topic of Section 6.

5. Proof of the Main Theorem

In the ensuing discussion, the following quantities will prove useful. Let

$$\gamma(\alpha, x) \equiv \int_0^x e^{-t} t^{\alpha-1}\, dt, \qquad x \geq 0,\ \operatorname{Re} \alpha > 0, \tag{5.1}$$

denote the *incomplete gamma function* with parameter α. Furthermore, let

$$ {}_2F_1(a, b; c; \xi) \equiv \sum_{k=0}^{\infty} \frac{1}{k!} \frac{(a)_k (b)_k}{(c)_k} \xi^k, \tag{5.2} $$

denote the *hypergeometric* function where $(a)_k \equiv (a + k - 1), \ldots, (a - 1)\, a$ is *Pochammer's symbol.* We refrain from discussing the domain of convergence of the series above except to note that it is convergent when the argument has absolute value less than unity. The reader is referred to the standard references (Gradshteyn and Ryzhik, 1980; Magnus *et al.,* 1966; Mehta, 1967). From (Gradshteyn and Ryzhik, 1980, p. 663), we find the formula

$$ \int_0^\infty x^{u-1} e^{-bx} \gamma(v, ax)\, dx = \frac{a^v \Gamma(u + v)}{v(a + b)^{u+v}} \frac{a}{a + b} $$

$$ [\text{Re } (a + b) > 0, \text{ Re } (b) > 0, \text{ Re } (u + v) > 0]. \tag{5.3} $$

We will also need the formulas (Gradshteyn and Ryzhik, pp. 1042–1043)

$$ {}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)}, \qquad \text{Re } (c - a - b) > 0, \tag{5.4} $$

and

$$ {}_2F_1(a, b; c; \xi) = (1 - \xi)^{c-a-b}\, {}_2F_1(c - a, c - b; c; \xi). \tag{5.5} $$

Finally, we have the integral representations

$$ {}_2F_1(a, b; c; \xi) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c - b)} \int_0^1 t^{b-1}(1 - t)^{c-b-1}(1 - t\xi)^{-a}\, dt $$

$$ [\text{Re } c > 0, \text{ Re } b > 0, \mid \arg (1 - \xi) < \pi] \tag{5.6} $$

and

$$ \frac{x^p}{p}\, {}_2F_1(p, 1 - q; p + 1; x) = B(p, q; x) \qquad [0 \le x \le 1, p > 0, q > 0], \tag{5.7} $$

where

$$ B(p, q; x) \equiv \int_0^x t^{p-1}(1 - t)^{q-1}\, dt \tag{5.8} $$

is the *incomplete beta function.* One has the well-known equality

$$B(p, q, 1) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \qquad p, q, >0. \tag{5.9}$$

Equations (5.6) and (5.7) may be found in (Magnus *et al.*, 1966, pp. 54 and 950, respectively). Thus armed, we proceed.

Elementary calculations and a simple induction which we will not reproduce show that if $\{\mathbf{X}_{ij}, 1 \leq i, j \leq n\}$ is a family of nonnegative i.i.d. random variables with common density

$$f(x) = \int_0^\infty f_{\alpha,r+1}(x)\, d\mu(r), \qquad x \geq 0 \tag{5.10}$$

(see Eqs. (4.2)–(4.6), then for c > 0, $(1/c)X_{ij}$ has density

$$\int_0^\infty f_{c\alpha,r+1}(x)\, d\mu(r), \qquad x \geq 0,$$

while $\sum_{k\neq j} X_{ik}$ has density

$$\int_0^\infty f_{\alpha,r+n-1}(x)\, d\mu^{(n-1)}(r), \qquad x \geq 0,\ n \geq 2,$$

where $\mu^{(n-1)}$ denotes the $(n-1)$-fold convolution of μ with itself. With P_{ij} as in (4.1), we obtain for $0 < a < 1$,

$$\begin{aligned}
&\Pr\{P_{ij} \geq a\} \\
&= \Pr\left\{\frac{(1-a)}{a} X_{ij} \geq \sum_{k\neq j} X_{ij}\right\} \\
&= \int_0^\infty d\mu(r) \int_0^\infty d\mu^{(n-1)}(s) \int_0^\infty dx \int_0^x dy\, f_{a\alpha(1-a),\, r+1}(x) f_{\alpha,s+n-1}(y) \\
&= \int_0^\infty d\mu(r) \int_0^\infty d\mu^{(n-1)}(s) \frac{\Gamma(r+s+n)}{\Gamma(r+1)\Gamma(s+n-1)} \frac{(1-a)^{(s+n-1)}}{(s+n-1)} \\
&\qquad\qquad {}_2F_1(s+n-1; -r; s+n; 1-a) \\
&= \int_0^\infty d\mu(r) \int_0^\infty d\mu^{(n-1)}(s) \frac{\Gamma(r+s+n)}{\Gamma(r+1)\Gamma(s+n-1)} \\
&\qquad\qquad B(s+n-1, r+1; 1-a) \\
&= \int_0^\infty d\mu(r) \int_0^\infty d\mu^{(n-1)}(s) \frac{\Gamma(r+s+n)}{\Gamma(r+1)\Gamma(s+n-1)}
\end{aligned}$$

$$\int_0^{1-a} t^{s+n-2}\,(1-t)^r\,dt$$

$$= \int_0^\infty d\mu(r) \int_0^\infty d\mu^{(n-1)}(s)\, \frac{\Gamma(r+s+n)}{\Gamma(r+1)\Gamma(s+n-1)}$$

$$(1-a)^{s+n-1} \int_0^1 t^{s+n-2}\,(1-t+at)^r\,dt$$

using Fubini's theorem and Eqs. (5.1) through (5.8). Using (5.9), we see that this latter formula holds for $a = 0, 1$ as well. Since r and a are positive in the last integrand above, (2.7) and (2.9) imply the fundamental inequality

$$\Pr\{P_{ij} \geq a\} \geq [(1-a)G_\mu(1-a)]^{(n-1)}. \tag{5.11}$$

We remark that equality holds in (5.11) when μ is the unit mass concentrated at the origin. Then, as noted earlier at the outset of Section 4, the P_{ij} are distributed as the interval lengths obtained by uniform partitions of the unit interval. For this case, $G_\mu \equiv 1$ and the right-hand side of (5.11) yields $(1-a)^{n-1}$ as is well known.

Now since $G_\mu(\cdot)$ is smooth on the interval (0, 1) and $\lim_{x\to 1} G'_\mu(x) = \bar{\mu} < \infty$, a brief exercise in elementary calculus shows that there exists a constant $\eta > 0$ *dependent on* μ such that if $1 - \eta \leq x \leq 1$, then $G_\mu(x) \geq x^{\bar{\mu}+1}$. Combining this with (5.11), we obtain

$$\Pr\{P_{ij} \geq a\} \geq (1-a)^{(\bar{\mu}+2)(n-1)}, \qquad 0 \leq a \leq \eta. \tag{5.12}$$

We now turn to the case where the density f is the χ-squared density with one degree of freedom, i.e.,

$$f(x) = \frac{1}{\sqrt{(2\pi x)}} \exp\left(-\frac{x}{2}\right), \qquad x > 0.$$

The steps leading to (5.12) may be repeated to obtain

$$\Pr\{P_{ij} \geq a\} = (1-a)^{(n-1)/2}\, H_n(a), \tag{5.13}$$

where

$$H_n(a) \equiv \frac{\Gamma(n/2)}{\Gamma(\frac{1}{2})\Gamma((n+1)/2)}\, {}_2F_1\left(\frac{n-1}{2}, \frac{1}{2}; \frac{n+1}{2}; 1-a\right) \tag{5.14a}$$

$$= \frac{1}{\pi} \int_0^1 t^{-1/2}\,(1-t)^{n/2-1}\,(1-t-at)^{-n/2+1/2}\,dt \tag{5.14b}$$

$$= \frac{2}{\pi} \int_0^{\pi/2} \frac{d\theta}{(1 + a \tan^2\theta)^{(n-1)/2}} \quad 0 \le a \le 1 \tag{5.14c}$$

using $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, Eq. (5.6), and the change of variables $t = \sin^2 \theta$, $0 \le \theta \le \pi/2$.

Since the map $x \to x^{(n-1)/2}$ is convex on [0, 1], *Jensen's* inequality (see, e.g., Rudin, 1974) implies that

$$H_n(a) \geq (H_3(a))^{(n-1)/2}. \tag{5.15}$$

Using the change of variable $u = \tan \theta$ and the decomposition of the resulting integrand into partial fractions, we obtain

$$H_3(a) = \left(\frac{1 - a^{1/2}}{1 - a}\right). \tag{5.16}$$

Thus in this case, we have from (5.13), (5.15), and (5.16)

$$\Pr\{P_{ij} \geq a\} \geq (1 - a^{1/2})^{(n-1)/2}. \tag{5.17}$$

We have established the following key,

5.1. LEMMA. *Let P_{ij} be as in the assertion of the Main Theorem. Then*

(a) *For $f \in F_\alpha$, there exists a constant $C \geq 1$ such that*

$$\Pr\{P_{ij} \geq a\} \geq (1 - a)^{C(n-1)}. \tag{5.18a}$$

(b) *For f the χ-squared density with one degree of freedom,*

$$\Pr\{P_{ij} \geq a\} \geq (1 - a^{1/2})^{(n-1)/2}. \tag{5.18b}$$

The assertion of the Main Theorem may now be established in a relatively straightforward fashion. We prove it for the case where $f \in F_\alpha$. The steps for the remaining case are similar and are left to the reader. Set $K \equiv \lg C$. Then

$$\begin{aligned} &\Pr\{T(k, \mathbf{P}, n) > \lg (k + 2) + (2 + \delta) \lg (n) + K\} \\ &\geq \Pr\{\tau^{C(k+2)n^{2+\delta}} < 2^{-(k+2)}\} \\ &= \Pr\{\tau \leq 2^{-1/Cn^{2+\delta}}\} \end{aligned}$$

$$= \Pr\left\{\frac{1}{2}\max_{i<j}\sum_{k=1}^{n}|P_{ik} - P_{jk}| \le 2^{-1/Cn^{2+\delta}}\right\}$$

$$= \Pr\left\{\frac{1}{2}\max_{i<j}\sum_{k=1}^{n}(P_{ik} + P_{jk}| \le 2^{-1/Cn^{2+\delta}}\right\}$$

$$= \Pr\left\{\frac{1}{2}\min_{i<j}\sum_{k=1}^{n}\min(P_{ik}, P_{jk} \le 2^{-1/Cn^{2+\delta}}\right\}$$

since the rows of $\mathbf{P}$ each sum to unity. Continuing, we have

$$\Pr\{T(k, \mathbf{P}, n) > \lg(k + 2) + (2 + \delta)\lg(n) + K\}$$

$$\ge \Pr\left\{\min_{i<j}\sum_{k=1}^{n}\min(P_{ik}, P_{jk}) \ge 1 - e^{-\log 2/Cn^{2+\delta}}\right\}$$

$$\ge \Pr\left\{\min_{i<j}\min(P_{i1}, P_{j1}) \ge 1 - e^{-\log 2/Cn^{2+\delta}}\right\}$$

$$= [\Pr\{P_{11} \ge 1 - e^{-\log 2/Cn^{2+\delta}}\}]^n$$

since the rows of $\mathbf{P}$ are i.i.d. random vectors. By Lemma 5.1, we see that there exists $N \ge 0$ such that $n > N$ implies

$$\begin{aligned}
&[\Pr\{P_{11} \ge 1 - e^{-\log 2/Cn^{2+\delta}}\}]^n \\
&\ge e^{-\log 2(n-1)/n^{1+\delta}} \\
&> e^{-\log 2/n^{\delta}} \\
&> 1 - \log 2/n^{\delta} \\
&> 1 - 1/n^{\delta},
\end{aligned}$$

and the proof of the Main Theorem is complete.

5.2. *Remark.* The constants above may all be explicitly computed, but depend on a detailed knowledge of the measure involved. The Main Theorem yields an overall asymptotic complexity of $O(n^3 \lg n)$ for the squaring algorithm. A starting point for establishing a lower bound on the running time of the algorithm is Corollary 2.7, but this appears to be difficult to do in general.

6. The Class of Densities F_α

The purpose of this section is to provide an alternate description of the class of densities introduced in Section 4. The ideas in the proof of the following theorem are all fairly standard, so that, for the sake of brevity, easy details are left to the reader to verify for himself. This should not prove burdensome to the reader acquainted with elementary Banach space theory. The result alluded to below is "intuitively obvious."

6.1. THEOREM. *For fixed $\alpha > 0$, the class F_α coincides with the set of elements in the (norm) closed convex hull* (c.c.h.) *in $L^1([0, \infty))$ of the set of gamma densities*

$$G_\alpha \equiv \left\{x \to \frac{\alpha^{r+1}x^r}{\Gamma(r+1)} \exp(-\alpha x),\ x \geq 0,\ r \geq 0\right\},$$

with finite mean. (Hence it is a dense subset of the c.c.h.)

Proof. Let H_α denote the convex hull of G_α above, $C_0([0, \infty))$ the space of continuous functions on $[0, \infty)$ vanishing at infinity with the *supremun* norm, and $L^\infty([0, \infty))$ the usual space of almost everywhere bounded measurable functions on $[0, \infty)$ with the same norm. Finally, let M denote the space of (regular) complex measures on $[0, \infty)$ (regularity is assured since this space is σ-compact), $P \subset M$ the set of probability measures on $[0, \infty)$, and P_0 the subset of P consisting of those measures with finite first moment. Finally, let D be the set of *delta masses* $\{\delta_t : t \geq 0\}$. The *Reisz Representation Theorem* (see, e.g., Rudin, 1974) exhibits M as the normed dual of $C_0([0, \infty))$ (M is endowed with the total-variation norm; see again Rudin, 1974), with P as its *state space*, i.e., the set of all positive linear functions of norm one. Endow P with the relative $\sigma(M, C_0([0, \infty)))$ topology, i.e., the weak*-topology. $L^\infty([0, \infty))$ is isometrically isomorphic to the normed dual of $L^1([0, \infty))$. Endow $L^1([0, \infty))$ with the $\sigma(L^1, L^\infty)$, i.e., the weak topology. If S is a subset of either M or $L^1([0, \infty)$, then $\overline{S}$ denotes its topological closure. Define the map $\Phi_\alpha : P \to L^1([0, \infty))$ via

$$\Phi_\alpha(\mu)(x) = \int_0^\infty \frac{\alpha^{r+1}x^r}{\Gamma(r+1)} \exp(-\alpha x)\, d\mu(r), \qquad x \geq 0. \tag{6.1}$$

The proof of the theorem rests on three simple facts:

(a) Φ_α is continuous with respect to the above topologies on the source and target,

(b) $\Phi_\alpha(P) \subset \overline{H_\alpha}$,

(c) $H_\alpha \subset \Phi_\alpha(P)$.

Assume that continuity has been established. The *Banach–Alaoglu* theorem (see, e.g., Rudin, 1973) shows that the unit ball in M is weak*-compact. $P \subset M$ is weak*-closed, hence weak*-compact. Since it is convex, the *Krein–Milman Theorem* (see, e.g., Rudin, 1973) implies that it is the weak*-c.c.h. of its set of extreme points, viz., D. Thus, one has

$$\Phi_\alpha(P) = \Phi_\alpha(\overline{D}) \subset \overline{\Phi_\alpha(D)} = \overline{H_\alpha},$$

establishing (b). The truth of (c) is self-evident. By the (assumed) continuity of Φ_α, we now observe that $\Phi_\alpha(P)$ is weakly compact, hence, weakly closed. Thus (b) and (c) together imply that

$$\Phi_\alpha(P) \subset \overline{H_\alpha} \subset \overline{\Phi_\alpha(P)} = \Phi_\alpha(P).$$

The characterization of $\Phi_\alpha(P_0)$ in the assertion of the theorem is a simple consequence of Fubini's theorem and the well-known relation

$$\lim_{x\to\infty} \frac{\Gamma(x+c)}{x^c\Gamma(x)} = 1, \qquad c \text{ real}, \tag{6.2}$$

which shows in fact that Φ_α and μ have the same orders of finite moments. We remark also that since H_α is convex, its norm and weak closures coincide.

Continuity of Φ_α is easily established. A basic open neighborhood of $\Phi_\alpha(\mu)$ in the weak topology is of the form

$$V = \left\{ f(\cdot) \in L^1 : \left| \int_0^\infty (f(x) - \Phi_\alpha(\mu)(x) g_i(x)\, dx \right| < \varepsilon, i = 1, 2, \ldots, m \right\} \tag{6.3}$$

for some positive ε and some integer m, with the functions $g_i(\cdot) \in L^\infty$. Define

$$h_i(r) = \int_0^\infty \frac{\alpha^{r+1}x^r}{\Gamma(r+1)} \exp(-\alpha x)\, dx, \qquad r \geq 0, i = 1, 2, \ldots, m. \tag{6.4}$$

A simple dominated convergence argument shows that (6.4) defines a continuous function. An application of Stirling's formula (4.7), together with another application of dominated convergence, similarly establishes that $h_i(\cdot)$ goes to zero at infinity. Therefore, the functions defined by (6.4) are functions in $C_0([0, \infty))$. Fubini's theorem now shows that the weakly open neighborhood of μ in P,

$$U = \left\{ v : \left| \int_0^\infty h_i(r)\, d\mu(r) - \int_0^\infty h_i(r)\, d\nu(r) \right| < \varepsilon, i = 1, 2, \ldots, m \right\}, \tag{6.5}$$

is mapped into the open set V of (6.3). This establishes the desired continuity and the proof is complete. ∎

We acknowledge Jim McKenna's assistance in bringing the following simple, but often overlooked, fact to our attention. This is the primary justification for the attention we have focused on the class F^{α}.

6.2. PROPOSITION. *The set* $\mathbf{F} = \bigcup_{\alpha>0} F_\alpha$ *is dense in the unit ball of the positive cone of* $L^1[0, \infty)$.

Proof. Let f be a positive probability density on $[0, \infty)$. Without loss of generality, we may take f to be continuous, of compact support. We form

$$f_m(x) = \sum_{r=0}^{\infty} \int_{r/m}^{(r+1)/m} f(y)\, dy \frac{(mx)^r}{\Gamma(r+1)} m \exp(-mx) \tag{6.6}$$

and also

$$g_m(x) = \sum_{r=0}^{\infty} f\left(\frac{k}{m}\right) \frac{(mx)^r}{\Gamma(r+1)} \exp(-mx). \tag{6.7}$$

An argument similar to that for Bernstein polynomials (see, for example, Feller, 1966, p. 222) shows that the g_m tend uniformly to f as $m \to \infty$. This convergence is also L^1 as the L^1 norm of g_m is $O(\|f\|_1)$. The function f_m lies in the unit ball and, for large m, approximates g_m arbitrarily well in the L^1 topology. The details may be easily supplied by the reader. ∎

Let $\hat{f}(\zeta) \equiv \int_0^\infty e^{-sx} f(x)\, dx$ be the *Laplace transform* of the density f, ζ a complex variable. Then one has the elementary

6.3. PROPOSITION. $\hat{f}$ *is analytic in the complement of the closed set* $(-\infty, -2\alpha] \cup \{\zeta : |\zeta + \alpha| \le \alpha\}$.

Proof. This is immediate on realizing that

$$\hat{f}(\zeta) = G_\mu\left(\frac{\alpha}{\alpha + \zeta}\right).$$

Use Fubini theorem for Re $\zeta > 0$ and analytic continuation for the rest. ∎

Conversely, given a function ψ analytic in the above region, one may express it in the form $\psi(\zeta) = G(\alpha/(\alpha + \zeta))$ for G analytic in the cut disk Δ of (1.7). The natural question which arises is when such a G is the generation function of a probability measure μ on $[0, \infty)$. An answer to this question coupled with the unicity of the Laplace transform would yield another characterization of the class F_α. For μ supported on the integers, the answer is well known (see Feller, 1966, p. 221). For the case of general μ, we are not aware of the answer.

7. Conclusion

In this note, we have examined the behavior of the power method for what might be considered *dense* stochastic matrices. Our analysis shows that for this class and for an interesting class of measures on the space of problem instances, the method performs reasonably well. The term *power* method is more often applied to the procedure which successively computes the iterates $\mathbf{v}(i + 1)' = \mathbf{v}(i)' \cdot \mathbf{P}$, where $\mathbf{v}(0)'$ is a fixed positive row-vector. Our results imply an asymptotic, probabilistically sharp running time estimate of $O(n^4)$ for this procedure, compared to the $O(n^3 \lg n)$ bound for the squaring method. We thank Albert Greenberg for pointing out that one may also derive a simple stopping criterion for the power method, from an elementary proof of the existence of the limiting vector. The estimation of the running time of the simplest implementation of this criterion, however, involves the distribution of the smallest entry of the initial stochastic matrix. One may proceed to derive this in the spirit of the calculations made here, but as our objective was to establish direct comparison with the result of *Kostlan* (1985), we have refrained from doing so. The results should nonetheless be of the same low dimensional polynomial complexity.

A problem which we have not considered here is the behavior of the method in the case of sparse stochastic matrices. One expects an asymptotically tight bound of $O(\lg n)$ here as well, but this class of matrices has been assigned measure zero by our model. It is conceivable that the error estimate may be tighter in this case. This problem is still open, even for iterations to find subsets of the spectrum of symmetric and Hermitian matrices. Here the chief problem (as always) is finding an acceptable measure on the space of problem instances that will reasonably reflect the kinds of matrices one is likely to encounter in practice. There are two obvious approches here. The first is to work with a fixed underlying incidence graph and to consider matrices with this fixed structure. The second (more difficult, but also more interesting mathematically) is to consider random models of the underlying graph structure along with the distributions of the nonzero entries.

Finally, we note that the usual objection to squaring matrices, viz., that it tends to cause *fill-in* of sparse matrices, does not apply here, as the limiting form of the powers of the stochastic matrix is one in which every entry is strictly positive. Naturally, however, if one could find a faster indirect method which obviated the need to retain the powers of the matrix, this would be highly preferable, from the point of view of efficient storage management. The power method is but one of many possible algorithms which may be used for eigenvalue problems. A worthwhile enterprise would be to carry out the same kind of analysis as done here for other standard routines such as *inverse iteration with shifts*, to name just one of the many other candidates.

APPENDIX

In this section, deferred proof of assertions in the main text are given.

Proof of Lemma 2.4. (a) By definition we have

$$\pi_k = \sum_{j=1}^{n} P_{jk}\pi_j \qquad \text{and} \qquad \sum_{j=1}^{n} \pi_j = 1.$$

Thus,

$$P_{ik} - \pi_k = \sum_{j=1}^{n} (P_{ik} - P_{jk})\pi_j$$

so that

$$\frac{1}{2}\sum_{k=1}^{n} |P_{ik} - \pi_k| \leq \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} |P_{ik} - P_{jk}|\pi_j \leq \tau(\mathbf{P}).$$

The other side of the inequality in (a) follows immediately from the usual triangle inequality.

(b) For $\mathbf{x} > 0$ we have

$$(\mathbf{Px})_i - (\mathbf{Px})_j = \sum_{k=1}^{n} (P_{ik} - P_{jk})x_k.$$

Let $A_{ij} \equiv \{k : P_{ik} > P_{jk}\}$. Then since all row sums are unity,

$$\sum_{k\in A_{ij}} (P_{ik} - P_{jk}) = \sum_{k\in A_{ij}^c} (P_{jk} - P_{jk}).$$

Therefore,

$$\frac{1}{2}\sum_{k=1}^{n} |P_{ik} - P_{jk}| = \sum_{k\in A_{ij}} (P_{ik} - P_{jk})$$
$$= \sum_{k\in A_{ij}^c} (P_{jk} - P_{jk})$$

so that we have

$$(\mathbf{Px})_i - (\mathbf{Px})_j = \sum_{k\in A_{ij}} (P_{ik} - P_{jk})x_k - \sum_{k\in A_{ij}^c} (P_{jk} - P_{ik})x_k$$
$$\geq \frac{1}{2}\sum_{k=1}^{n} |P_{ik} - P_{jk}|\sigma(\mathbf{x}).$$

This shows that,

$$\tau(\mathbf{P}) \geq \sup\left\{\frac{\sigma(\mathbf{Px})}{\sigma(\mathbf{x})} : \mathbf{x} > \mathbf{0},\ \sigma(\mathbf{x}) \neq \mathbf{0}\right\}.$$

In the other direction, given a subset $S \subset \{1, 2, \ldots, n\}$, we let $\mathbf{1}_S$ denote the column-vector with,

$$(\mathbf{1}_S)_i = \begin{cases} 1, & \text{if } i \in S, \\ 0, & \text{otherwise.} \end{cases}$$

Choose indices i and j such that,

$$\tau(\mathbf{P}) = \frac{1}{2}\sum_{k=1}^{n} |P_{ik} - P_{jk}|$$
$$= \sum_{k\in A_{ij}} (P_{ik} - P_{jk}).$$

Without loss of generality, $A_{ij} \neq \varnothing$ as the statement of the lemma is otherwise trivially true. Also since each row sum is unity, $A_{ij} \neq \{1, 2, \ldots, n\}$. Continuing, we see that,

$$\tau(\mathbf{P}) = (\mathbf{P1}_{A_{ij}})_i(\mathbf{P1}_{A_{ij}})_j$$
$$\leq \sigma(\mathbf{P1})_{A_{ij}})$$
$$= \frac{\sigma(\mathbf{P}(\mathbf{1}_{A_{ij}} + \mathbf{1})}{\sigma(\mathbf{1}_{A_{ij}} + \mathbf{1})}$$

so that the supremum in (b) is actually attained.

(c) This is an immediate consequence of (b) as

$$\begin{aligned}\tau(\mathbf{P}_1\mathbf{P}_2) &= \sup\left\{\frac{\sigma(\mathbf{P}_1\mathbf{P}_2\,\mathbf{x})}{\sigma(\mathbf{x})} : \mathbf{x} > \mathbf{0},\ \sigma(\mathbf{x}) \neq \mathbf{0}\right\} \\ &\leq \tau(\mathbf{P}_1) \sup\left\{\frac{\sigma(\mathbf{P}_2\mathbf{x})}{\sigma(\mathbf{x})} : \mathbf{x} > \mathbf{0},\ \sigma(\mathbf{x}) \neq \mathbf{0}\right\} \\ &= \tau(\mathbf{P}_1)\tau(\mathbf{P}_2).\end{aligned}$$

(d) For this we note that

$$\begin{aligned}\|\mathbf{P}^{k+1} - \Lambda(\pi)\| &= \|\mathbf{P}(\mathbf{P}^k - \Lambda(\pi)\| \\ &\leq \|\mathbf{P}\|\,\|\mathbf{P}^k - \Lambda(\pi)\| \\ &= \|\mathbf{P}^k - \Lambda(\pi)\|.\end{aligned}$$

This completes the proof. ■

Proof of Proposition 2.6. This amounts to little more than justifying differentiation under the integral sign. For fixed $\xi \in \Delta \setminus (-1, 0]$, we have

$$\left| \int_0^\infty r\xi^{r-1}\, d\mu(r) \right| \leq \int_0^\infty \left| \frac{(\xi + h)^r - \xi^r}{h} - r\xi^{r-1} \right| d\mu(r).$$

The integrand is bounded by an integrable function of r as $h \to 0$ (use Stirling's formula (4.7)). Dominated convergence yields the result for the first derivative and higher derivatives follow similarly by induction. ■

Proof of Lemma 3.1. (a) If $\|\mathbf{P} - \Lambda(\pi)\| \leq \varepsilon$, then

$$\begin{aligned}\|\mathbf{x} - \pi\| &= \sum_{k=1}^n \left| \frac{1}{n} \sum_{j=1}^n P_{jk} - \pi_k \right| \\ &\leq \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n |P_{jk} - \pi_k| \\ &\leq \varepsilon\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{P} - \Lambda(\pi)\| &\leq \|\mathbf{P} - \Lambda(\mathbf{x})\| + \|\Lambda(\mathbf{x}) - \Lambda(\pi)\| \\ &\leq 2\varepsilon.\end{aligned}$$

(b) Let $\|\mathbf{P} - \Lambda(\mathbf{x})\| \leq \varepsilon$. A simple induction shows that $\|\mathbf{P}^k - \Lambda(\pi)\| \leq \varepsilon$, $k \geq 1$. Passage to the limit in k yields the desired result. To obtain the other inequality, use the triangle inequality as before.

Proof for Example 4.3(b). Let

$$\mu_{2k+\rho-1} = \frac{\rho}{2^{2k+\rho}} \frac{\Gamma(2k+\rho)}{\Gamma(k+\rho+1)} \frac{1}{k!}$$

as in (4.10). Then

$$\begin{aligned}\sum_{k=0}^{\infty} \mu_{2k+\rho-1} &= \sum_{k=0}^{\infty} \frac{\rho}{2^{2k+\rho}} \frac{\Gamma(2k+\rho)}{\Gamma(k+\rho+1)} \frac{1}{k!} \\ &= \frac{\rho}{2\sqrt{(\pi)}} \sum_{k=0}^{\infty} \frac{\Gamma(k+\rho/2)\Gamma(k+\rho/2+\frac{1}{2})}{\Gamma(k+\rho+1)k!},\end{aligned}$$

and using the duplication formula for the Γ function (see, Magnus *et al.*, 1966, p. 3),

$$\Gamma(2\xi) = \frac{1}{\sqrt{\pi}} 2^{2\xi-1}\Gamma(\xi)\Gamma\left(\xi + \frac{1}{2}\right).$$

Continuing, we have

$$\begin{aligned}\sum_{k=0}^{\infty} \mu_{2k+\rho-1} &= \frac{\rho}{2\sqrt{\pi}} {}_2F_1\left(\frac{\rho}{2}, \frac{\rho+1}{2}; \rho+1; 1\right) \frac{\Gamma(\rho/2)\Gamma(\rho+1)/2)}{\Gamma(\rho+1)} \\ &= \frac{\rho}{2\sqrt{\pi}} \frac{\Gamma(\rho+1)\Gamma(\frac{1}{2})}{\Gamma(\rho/2+1)\Gamma((\rho+1)/2)} \frac{\Gamma(\rho/2)\Gamma(\rho+1)/2)}{\Gamma(\rho+1)} \\ &= \frac{(\rho/2)\Gamma(\rho/2)}{\Gamma(\rho/2+1)} \frac{\Gamma(\frac{1}{2})}{\sqrt{\pi}} \\ &= 1\end{aligned}$$

since $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. ■

Acknowledgments

The author expresses his appreciation to Professor Steven Smale, under whose direction this project was undertaken, as well as to Professor F. Alberto Grünbaum, Professor Richard Karp, and Professor Michael Sipser, who reviewed the manuscript. He also thanks

Professor David Aldous and Eric Kostlan for early helpful discussions. It also gives him great pleasure to thank Albert Greenberg and Jim McKenna (Bell Labs) for their valuable assistance.

REFERENCES

ALDOUS, D. (1983), On the time taken by walks on finite groups to visit every state, *Z. Wahrsch. Geb.* **62,** 361–374.

ALDOUS, D. (1982), Random walks on finite groups and rapidly mixing Markov chains, *Sém. Probab. XVII.*

FELLER, W. (1966), "An Introduction to Probability Theory and Its Applications," Vol. 2, Wiley, New York.

GRADSHTEYN, I., AND RYZHIK, M. (1980), "Tables of Integrals, Series and Products," Academic Press, New York.

KOSTLAN, E. (1985), "Statistical Complexity of Numerical Linear Algebra," Thesis, University of California, Berkeley.

LANDAU, H. J., AND ODLYZKO, A. M. (1981), Bounds for eigenvalues of certain stochastic matrices, *Linear Algebra and Appl.* **38,** 5–15.

MAGNUS, W., OBERHETTINGER, F., AND SONI, R. P. (1966), "Formulas and Theorems for the Special Functions of Mathematical Physics," Springer-Verlag, New York.

MEHTA, M. (1967), "Random Matrices and the Statistical Theory of Energy Levels," Academic Press, New York.

RUDIN, W. (1974), "Real and Complex Analysis," McGraw–Hill, New York.

RUDIN, W. (1973), "Functional Analysis," McGraw–Hill, New York.

SENETA, E. (1980), "Non-negative Matrices and Markov Chains," Springer-Verlag, New York/Heidelberg/Berlin.

SHUB, M., AND SMALE, S. (1985), Computational complexity: On the geometry of polynomials, and a theory of cost, Part 1, *Ann. Sci. École Norm. Sup. (4).*

SHUB, M., AND SMALE, S. (1986), Computational complexity: On the geometry of polynomials and a theory of cost, Part 2, *Siam J Comp.* **15,** No. 1, 145–161.

SMALE, S. (1981). The fundamental theorem of algebra and complexity theory, *Bull. Amer. Math. Soc. (NS)* **4,** 1–36.

SMALE, S. (1983a), On the average number of steps in the Simplex method of linear programming, *Math. Prog.* **27,** 241–262.

Smale, S. (1983b), The problem of the average speed of the Simplex Method, in "Mathematical Programming: The State of the Art," Bonn 1982 (A. Bachem *et al.*, Eds.), Springer-Verlag, Berlin/New York.

SMALE, S. (1985), On the efficiency of the algorithms of analysis, *Bull. Amer. Math. Soc. (NS)* **13,** 87–121.

WHITTAKER, E. T., AND WATSON, G. N. (1962), "A Course of Modern Analysis," Cambridge Univ. Press, London/New York.